CDF note 7174

Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using Lepton Plus Jets Events with Soft Muon b-Tagging

The CDF Collaboration
URL http://www-cdf.fnal.gov/physics/new/top/public/ljets/slt/public_200.html
(Dated: August 7, 2004)

We present a measurement of the $t\bar{t}$ production cross section in $\sim 200\,\mathrm{pb}^{-1}$ of Run 2 data using events with a lepton, 3 or more jets, and missing E_T. Events consistent with $t\bar{t}$ decay are found by identifying jets containing candidate heavy-flavor semileptonic decays to muons. Backgrounds are computed from a combination of Run 2 data and simulation. Signal acceptance is determined from Run 2 `PYTHIA` Monte Carlo. Based on the tags in events with 3 or more jets, a production cross section of $4.2^{+2.9}_{-1.9} \pm 1.4\,\mathrm{pb}$ is measured.

Preliminary Results for Summer 2004 Conferences
Contact e-mail: anyes@uiuc.edu, lcerrito@uiuc.edu, grundler@uiuc.edu, mendi@uiuc.edu, tml@uiuc.edu

I. INTRODUCTION

This note describes a measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF detector at the Fermilab Tevatron. The standard model predicts that $t\bar{t}$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV proceeds primarily through quark-antiquark annihilation with a small admixture from gluon fusion. The cross section is calculated to be $6.7^{+0.7}_{-0.9}$pb at a top mass of 175 GeV/c^2 [1]. The calculated cross section decreases by approximately 0.2pb for each 1 GeV/c^2 increase in the top mass over the range 170<M_{top} <190 GeV/c^2. This measurement uses muon tagging for b-jet identification in order to reduce the background from W plus multijet production. The measurement of the $t\bar{t}$ production cross section provides a test of the QCD calculations, and a significant deviation from the predicted cross section could signal beyond the standard model production mechanisms for $t\bar{t}$ pairs.

The CDF detector is described in detail in [2].

II. DATA SAMPLE & EVENT SELECTION

This analysis is based on an integrated luminosity of 194 pb^{-1} collected with the CDFII detector between March 2002 and August 2003. The data are collected with an inclusive lepton trigger that requires an electron (muon) with E_{T} >18 GeV (P_{T} >18 GeV/c). From this inclusive lepton dataset we select events offline with a reconstructed isolated electron E_{T} (muon P_{T}) greater than 20 GeV, missing E_{T} >20 GeV and at least 3 jets with E_{T} >15 GeV.

The dataset selected above, called "lepton+jets", is dominated by QCD production of W bosons with multiple jets. As a first stage of background reduction, we define a total event energy, H_{T}, as the scalar sum of the electron E_{T}, muon P_{T}, missing E_{T} and jet E_{T} for jets with E_T >8 GeV and $|\eta|$<2.4. Figure 1 shows $S/\sqrt{S+B}$ as a function of H_{T} and demonstrates that an H_{T} cut improves the sensitivity of the measurement. We find H_T >200 GeV to be the optimal choice. Requiring H_T >200 GeV rejects approximately 40% of the background while retaining more than 95% of the $t\bar{t}$ signal.

Even after the H_{T} cut, the expected S/B in $W+$ 3 or more jet events is only of order 1:1. To further improve the signal to background we identify events with one or more b-jets by searching for semileptonic decays of B hadrons into muons inside jets. This technique is called soft lepton tagging, or SLT.

In what follows, we refer to the $W+$ 3 or more jet sample after requiring H_T >200 GeV, but before requiring a soft lepton tag, as the "pre-tag" sample. In 194 pb^{-1} we find 337 pre-tag events, 115 from $W \to \mu\nu$ and 222 from $W \to e\nu$.

A. Soft Lepton Tagging Algorithm

Muon identification at CDF proceeds by matching extrapolated tracks found in the central tracker to track segments reconstructed in the muon chamber. Matching is done in extrapolated position in the muon chamber drift direction and, where such information is available, in the coordinate along the chamber wires, and in the extrapolated slope compared to the slope of the reconstructed muon chamber track segment. The matching distributions, between the measured muon chamber hits and the extrapolated track, are a function of P_{T}, η and ϕ. The muon SLT algorithm uses a global χ^2, L, built from the matching distributions, that separates muon candidates from background. In this analysis a jet is considered "tagged" if it contains an SLT muon with $P_{\mathrm{T}} > 3$ GeV/c, with L <3.5 and within $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$ <0.6 of the jet axis. The muon track is required to come within 5cm in Z (the beam direction) of the event primary vertex. The muon SLT algorithm does not use calorimeter information and so is efficient for muon identification inside jets. Two sets of muon drift chambers are used by the SLT, "CMUP" which covers the region $|\eta|< 0.6$ and "CMX", which covers the region $0.6 <|\eta|< 1.0$.

Events are rejected if the isolated high P_T lepton is a muon of opposite charge to an SLT muon tag that together with the SLT muon has an invariant mass consistent with a J/ψ and Υ a Z^0 or a sequential, double semi-leptonic $b \to c \to s$ decay.

B. Total $t\bar{t}$ Acceptance

1. *Geometric × Kinematic Acceptance*

The total acceptance is measured in a combination of data and Monte Carlo. The geometric times kinematic acceptance of the basic lepton+jets event selection is measured using the PYTHIA Monte Carlo program [3]. The

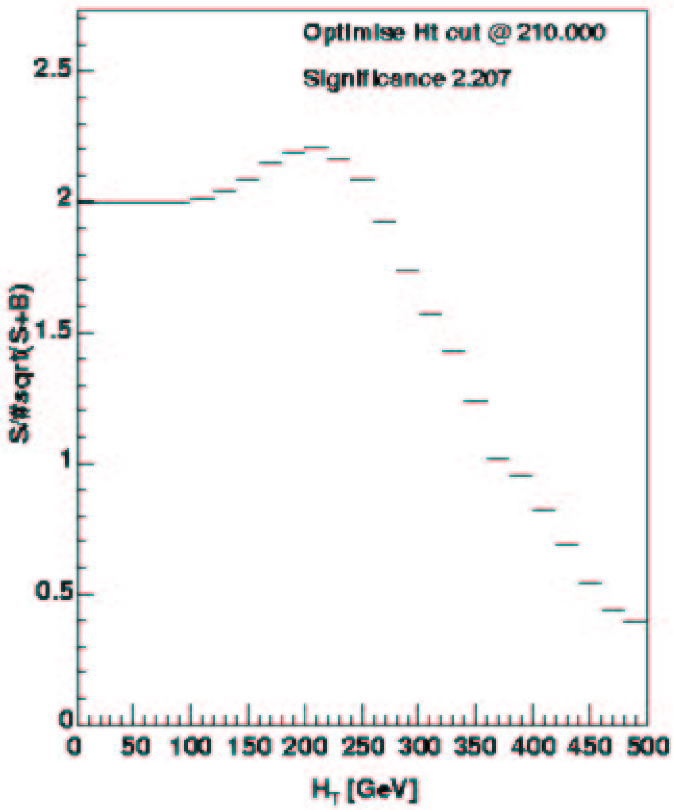

FIG. 1: Optimization of the H_{T} selection cut.

efficiency for identifying the isolated, high P_{T} lepton is scaled to the value measured in the data using the unbiased leg in Z-boson decays. The geometric times kinematic acceptance, as a function of the number of identified jets above 15 GeV, is shown in Table I. These numbers include the measured efficiencies of the high-P_{T} lepton triggers.

Jet multiplicity	1 jet	2 jets	3 jets	$\geq$ 4 jets
electrons	(0.204$\pm$ 0.005)%	(1.05$\pm$ 0.01)%	(1.79$\pm$ 0.02)%	(2.27$\pm$ 0.02)%
muons (CMUP)	(0.095$\pm$0.003)%	(0.50$\pm$0.007)%	(0.861$\pm$0.007)%	(1.12$\pm$ 0.1)%
muons (CMX)	(0.045$\pm$0.002)%	(0.235$\pm$0.006)%	(0.387$\pm$0.007)%	(0.507$\pm$ 0.008)%

TABLE I: Geometric times kinematic acceptance for $t\bar{t}$ events as a function of jet multiplicity from PYTHIA Monte Carlo, corrected for the data/MC ratio for tight lepton ID efficiencies and including trigger efficiencies. The uncertainties listed are statistical only.

2. *SLT Efficiency*

The muon identification efficiency of the SLT algorithm is measured in data using J/ψ Υ and Z^0 events. The measured efficiency vs. P_{T} for CMUP and for CMX are shown in Figures 2 and 3

The decrease in efficiency with increasing P_{T} is a result of non-Gaussian tails in the components of the global χ^2. The efficiency measurement is dominated by isolated muons whereas the muons in b-jets tend to be surrounded by other tracks. We have studied the dependence of the efficiency on N_{trk}, the number of tracks above 1 GeV/c in a cone of ΔR=0.4 around the muon track. We find no significant efficiency loss, although the precision of the measurement

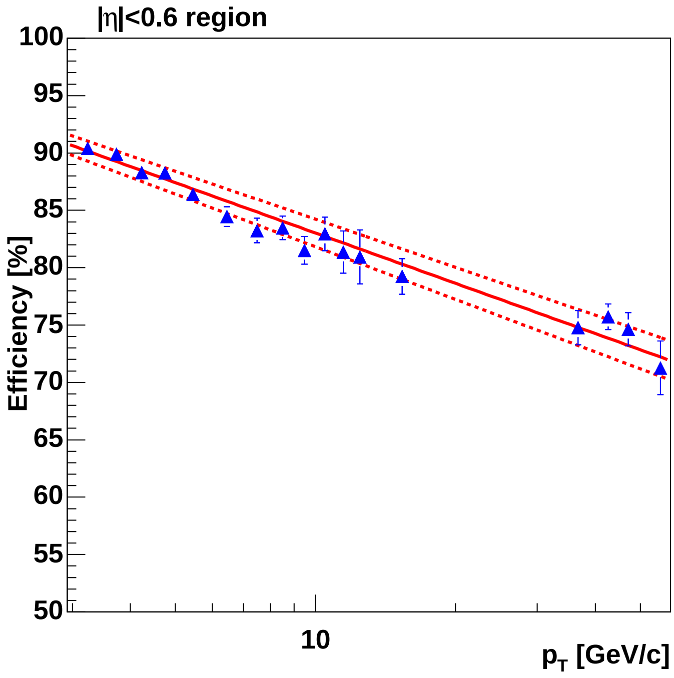

FIG. 2: The SLT efficiency for CMUP as a function of P_T as measured from J/ψ and Z^0 data for $|L|$ <3.5. The dotted lines are the 1σ uncertainties on the fit that are used in the evaluation of the systematic uncertainty.

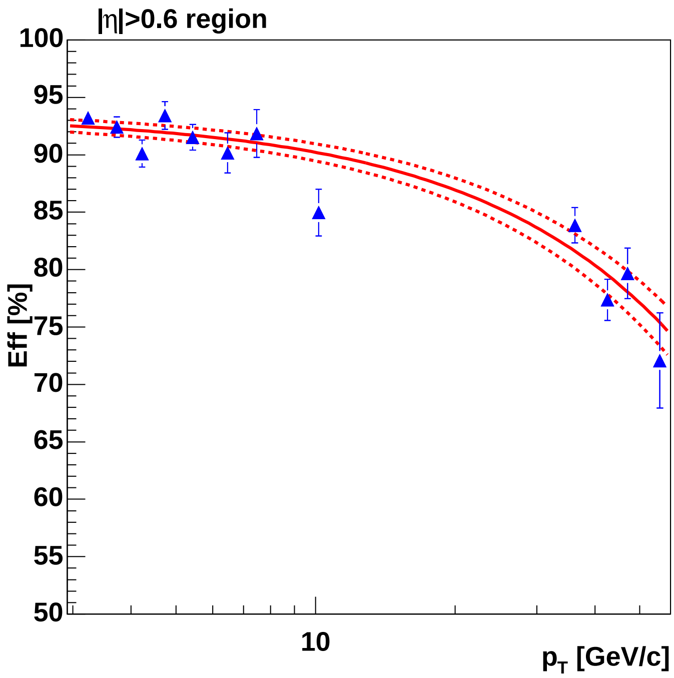

FIG. 3: The SLT efficiency for CMX as a function of P_T as measured from J/ψ and Z^0 data for $|L|$ <3.5. The dotted lines are the 1σ uncertainties on the fit that are used in the evaluation of the systematic uncertainty.

is poor near N_{trk}=6, the mean expected in $t\bar{t}$ events. We include a systematic error to account for this uncertainty by fitting the efficiency vs. N_{trk} to a linear function and evaluating this function at the mean N_{trk} expected in $t\bar{t}$ events. The systematic uncertainty on the efficiency for at least one SLT tag in a $t\bar{t}$ event from this effect is +0%, -8%.

The measured efficiencies shown in Figures 2 and 3 are applied directly in the Monte Carlo. The efficiency for finding one or more SLT tags in a $t\bar{t}$ event ("tagging efficiency") is shown in Table II. These efficiencies include mistags in $t\bar{t}$ events (i.e. tags that do not come from semileptonic heavy flavor decays), which contribute approximately 25% to the total tagging efficiency. The total $t\bar{t}$ detection efficiency is the product of the geometric×kinematic acceptance and the tagging efficiency.

Jet multiplicity	1 jet	2 jets	3 jets	$\geq$ 4 jets
electrons	$(8.8 \pm 0.7)\%$	$(13.1 \pm 0.4)\%$	$(14.6 \pm 0.3)\%$	$(15.7 \pm 0.3)\%$
muons	$(7.7 \pm 0.8)\%$	$(12.8 \pm 0.4)\%$	$(13.4 \pm 0.3)\%$	$(16.0 \pm 0.3)\%$

TABLE II: $t\bar{t}$ event tagging efficiency for SLT muons as a function of jet multiplicity from PYTHIA Monte Carlo. Uncertainties are statistical only.

III. BACKGROUNDS

The dominant background for this analysis is QCD production of W-boson plus multijet events. These events enter the signal sample when either one of the jets is a b-jet, or a light quark jet is mis-identified as containing a semileptonic B-hadron decay. We measure this background by constructing a "fake matrix" using jets in photon+jet events with one or more jets, as a function of P_{T}, η and ϕ. The fake matrix parameterizes the probability that a track with a given P_{T}, η and ϕ will satisfy the SLT requirement of L <3.5 . The fake probability is approximately 0.7% per taggable track (P_{T} >3 GeV/c, ΔR <0.6 from a jet axis and fiducial to the muon chambers). The background from QCD multijet events is calculated by summing the tagging probability for each taggable track in the lepton+jets sample, using the fake matrix and the P_{T}, η and ϕ of the track, and correcting for the probability of multiple tags in an event. The sum of the background contribution for all lepton+jet events is the total background due to tagged W plus jet events.

This technique for measuring the tagged W plus jets background relies on the assumption that the tagging rate in the jets in photon plus jet events is a good model for the tagging rate in the jets in W plus jet events. The assumption is plausible because the SLT tagging rate in generic jet events is dominated by fakes. We have studied the heavy-flavor content of the tags in the photon plus jets sample using the overlap between SLT tags and displaced vertex tags using the silicon tracker [4]. We find that 21.0±1.4% of the tags in the photon plus jets sample are from heavy flavor. Thus the background estimate is relatively insensitive to differences in the heavy flavor content of jets in photon plus jets events versus W plus jets events. The ideal place to test this assumption would be Z plus jet events, but the statistics there are limited. Instead we test the accuracy of the fake matrix for predicting SLT muon tags by using it to predict the number of tags in a variety of independent samples. We check Z plus jet events, events triggered on a jet with thresholds of 20, 50, 70 and 100 GeV and events triggered on the scalar sum of transverse energy in the detector. We find that the matrix predicts the observed number of tags in each of these samples to within 10%, as shown in Figure 4.

The other substantial background in this analysis comes from events without W bosons. These events are typically QCD jet events where one jet has faked a high-P_T lepton and mismeasured energies produce apparent missing E_{T}. We measure this "non-W" background by extrapolating the number of SLT tagged events with an isolated lepton and low missing E_{T} into the signal region of large missing E_{T}.

Residual Drell-Yan background is estimated from the data by extrapolating the number of events inside the Z-mass window that pass the selection cuts, including the SLT tag, to events in the signal region outside the Z-mass window. Other, small backgrounds from a variety of sources are estimated using the Monte Carlo.

The backgrounds as a function of jet multiplicity are summarized Table IV.

IV. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties in this analysis come from Monte Carlo modeling of the geometrical and kinematic acceptance, knowledge of the SLT tagging efficiency, the effect on the acceptance of the uncertainty on the jet energy scale, uncertainties on the background predictions, and the uncertainty on the luminosity.

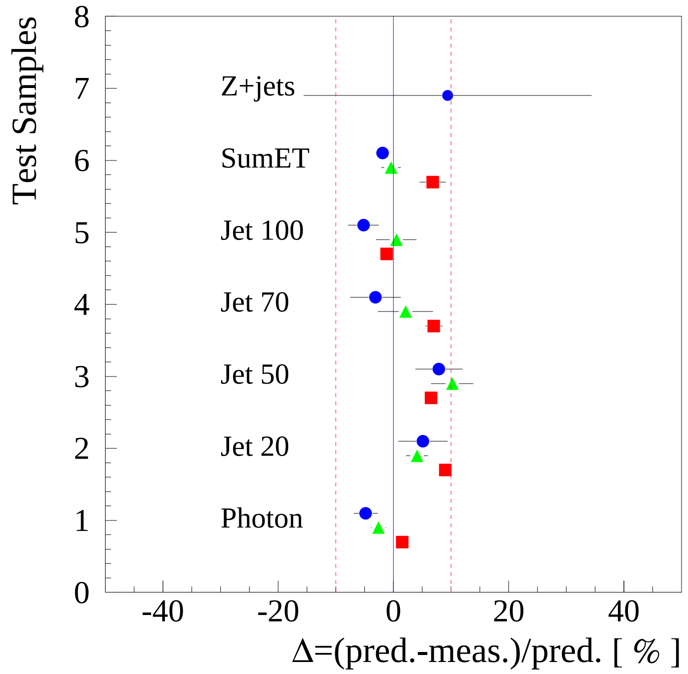

FIG. 4: The percent difference between the predicted and measured tags in a variety of samples. The red (squares)/green (triangles)/blue (circles) markers indicate there are 1/2/3-or-more fiducial jets in the event in addition to the trigger object (photon or jet).

Monte Carlo modeling of geometrical and kinematic acceptance include effects of PDFs, ISR and FSR, and jet energy scale. These are estimated by comparing different choices for PDFs, varying ISR, FSR and the jet energy scale in the Monte Carlo and comparing the `PYTHIA` generator with `HERWIG` [5]. The total systematic uncertainty due to these factors is 8.2% Modeling of the lepton ID efficiency in events with multiple jets is an additional source of systematic uncertainty on the acceptance. We use a data to Monte Carlo scale factor that is taken from inclusive Z data and Monte Carlo which is dominated by events with no jets. A 5% systematic uncertainty on this scale factor is estimated by convoluting the scale factor measured as a function of ΔR between the lepton and the nearest jet with the ΔR distribution of leptons in $\geq$3 jet $t\bar{t}$ events. Adding these uncertainties in quadrature gives a total Monte Carlo modeling systematic uncertainty of 9.6% .

There are several factors that contribute to the systematic uncertainty on the SLT tagging efficiency. The uncertainty due to the limited knowledge of the P_{T} dependence is determined by varying the efficiency curves used in the $t\bar{t}$ Monte Carlo for the tagging efficiency measurement according to the upper and lower bands in Figures 2 and 3. We find that the tagging efficiency for $t\bar{t}$ changes by $\pm$1% from its central value when varying the efficiency curves, and take this as a systematic uncertainty. An additional systematic uncertainty for the tagging efficiency comes from the fact that we implicitly use the Monte Carlo tracking efficiency for taggable tracks. As these tracks can be in dense environments in or near jets, we expect it to be less than 100%. Studies done by embedding Monte Carlo tracks in jets in data and jets in Monte Carlo events indicate that the Monte Carlo tracking efficiency in dense environments is a few percent higher than in data. We assign a 5% systematic uncertainty to the tagging efficiency for this effect. As described in Section II B 2 the systematic uncertainty due to the modeling of the isolation dependence of the tagging efficiency is +0%, -8%, based on extrapolating a linear fit of the efficiency loss vs. N_{trk} measured in J/ψ to the mean N_{trk} expected in $t\bar{t}$ events. Finally, the statistical uncertainty on the SLT tagging efficiency measurement, differences between `PYTHIA` and `HERWIG`, and estimation of the heavy flavor content of the mistag matrix, also come in as a systematic uncertainties. Adding these contributions in quadrature gives us an overall systematic uncertainty for the tagging efficiency of +7%, -10.7%.

Uncertainties on the fake matrix are determined by the level of agreement between observed tags and predictions in a variety of samples. This was described in Section III. We take 10% as the uncertainty on the fake+$Wb\bar{b}+Wc\bar{c}$

prediction.

Uncertainties on the non-W prediction are determined by studying the effect on the prediction when the missing $E_{\rm T}$ and isolation requirements are changed. A 50% systematic uncertainty is assigned to the non-W prediction and we add in quadrature to that the statistical uncertainty on the prediction in each jet bin. The total non-W systematic is 59% for $W \rightarrow e\nu$ events and 73% for $W \rightarrow \mu\nu$ events. In addition, we include a separate systematic uncertainty for the measurement of the non-W fraction in the pre-tag sample. This is also estimated by moving the missing $E_{\rm T}$ and isolation boundaries and added in quadrature with the statistical uncertainty on the non-W fraction, giving an uncertainty of 53% for electron events and 60% for muon events.

The systematic uncertainty on the small Drell-Yan background is dominated by its statistical uncertainty. Uncertainties on the Monte Carlo background predictions come from uncertainties in the cross sections for the various processes and from the statistics of the Monte Carlo samples.

The systematic uncertainties are summarized in Table III.

Source	Fractional Sys. Uncert.	Contribution to $\sigma_{t\bar{t}}$
Acceptance Modeling	9.6%	
SLT Tagging Efficiency	+7%,-11%	+19.3% -13.4%
Fake Matrix Prediction	10%	17.6%
Non-W Prediction	59%(e) 73%(μ)	19%
non-W Fraction	53%(e) 60%(μ)	13%
Drell-Yan and other MC backgrounds	19%	2.8%
Luminosity	6%	6%
Total Systematic Uncertainty		±34%

TABLE III: Summary of systematic uncertainties.

V. RESULTS

Table IV shows a summary of the background estimates for each jet bin and the number of SLT tagged events. The total background and the $t\bar{t}$ expectation are also listed. The line labeled "Corrected Background" includes an iterative correction to the background estimate. This correction is needed because we apply the fake matrix to the pre-tag events in order to estimate the Fakes+$Wb\bar{b}$+$Wc\bar{c}$. Since the events before tagging include some $t\bar{t}$, this results in an over estimate of the background for which we correct. Similarly the Wc background is proportional to the number of pre-tag events, and we correct for that as well.

Background	W + 1 jet	W + 2 jets	W + 3 jets	W +≥ 4 jets	W+≥ 3 jets
Events before tagging	18314	2889	226	111	337
Fake, Wbb, $Wc\bar{c}$	115.9 ± 11.6	41.2 ± 4.1	6.4 ± 0.6	4.3 ± 0.4	10.7 ±1.1
Wc	10.4 ± 2.9	4.1 ± 1.3	0.4 ± 0.1	0.12 ± 0.05	0.55 ± 0.18
WW, WZ, ZZ, Z→ $\tau^+\tau^-$	1.13± 0.17	1.36 ± 0.07	0.18 ± 0.03	0.04 ± 0.01	0.20 ± 0.02
non-W	21.1 ± 9.9	8.1 ± 3.9	1.5 ± 0.8	0.7 ± 0.5	2.4 ± 1.2
Drell-Yan	3.1 ± 0.6	0.64 ± 0.27	0.18 ± 0.14	0.0 ± 0.0	0.18 ± 0.14
Single-Top	0.51 ± 0.04	0.95 ± 0.06	0.15 ± 0.01	0.036 ± 0.003	0.19 ± 0.01
Total Background	152.2 ± 15.5	56.3 ± 5.9	8.9 ± 1.0	5.2 ± 0.7	14.2 ± 1.6
Corrected Background			11.59±1.5		11.59±1.5
$t\bar{t}$ expectation	0.36 ± 0.09	3.0 ± 0.5	5.6 ± 0.9	8.1 ± 1.8	13.7 ± 2.7
Total Background plus $t\bar{t}$	152.5 ± 15.5	59.3 ± 5.9	25.3±3.1		25.3 ± 3.1
Tagged Events	139	48	13	7	20

TABLE IV: Number of tagged events and the background summary. The $H_{\rm T}$ >200 GeV requirement is made only for events with at least 3 jets.

We calculate the cross section in the usual way as

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bckg}}{A_{t\bar{t}} \times \int \mathcal{L}dt} \tag{1}$$

where N_{obs} is the number of events with $\geq$ 3 tight jets that are tagged with at least 1 SLT, N_{bckg} is the corrected background and $A_{t\bar{t}}$ is the total acceptance (geometrical times kinematic times tagging efficiency), taken from Tables I and II For events with 3 or more jets, the total denominator is 2.01 $\pm$ 0.28 pb^{-1}.

For $t\bar{t}$ events in 3 or more jets, we find a cross section of

$$4.2^{+2.9}_{-1.9} \pm 1.4 \text{ pb},$$

where the first uncertainty is statistical and the second is systematic.

Figure 5 shows the number of tags in $W+1, 2, \geq 3$ jet events together with a histogram representing the components of the background. Figure 6 shows the same distribution with the $t\bar{t}$ contribution to each jet bin, normalized to the theoretical cross section of 6.7pb.

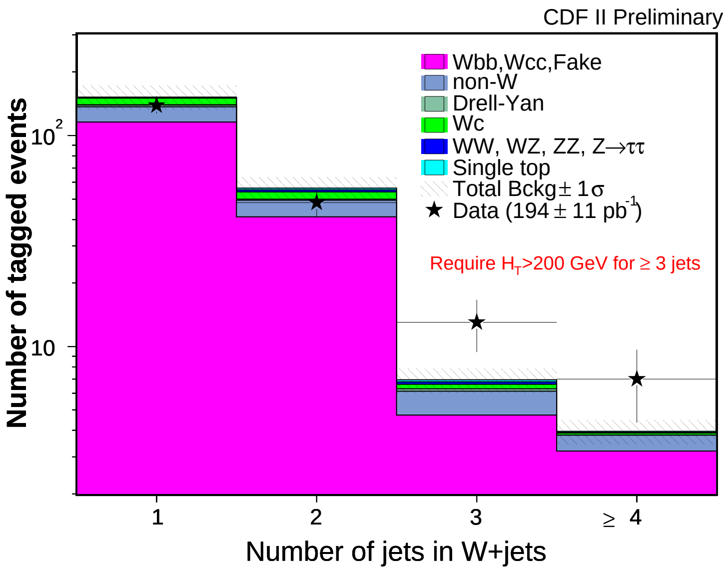

FIG. 5: The expected background and observed tags in $W+$ 1, 2, 3 and 4 or more jet events. The background is corrected for the $t\bar{t}$ content of the pretagged sample.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium fuer Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comision Interministerial de Ciencia y Tecnologia, Spain; in part by the European Community's Human Potential Programme under contract HPRN-CT-2002-00292, Probe for New Physics; and by the Research Fund of Istanbul University Project No. 1755/21122001.

[1] M. Cacciari, et al., The $t\bar{t}$ Cross-section at 1.8 TeV and 1.96 TeV: A Study of the Systematics due to Parton Densities and Scale Dependence, JHEP **404**, 68 (2004).

[2] F. Abe, et al., Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988); D. Amidei, et al., Nucl. Instum. Methods Phys. Res. A **350**, 73 (1994); F. Abe, et al., Phys. Rev. D **52**, 4784 (1995); P. Azzi, et al., Nucl. Instrum. Methods Phys. Res. A **360**, 137 (1995); The CDFII Detector Technical Design Report, Fermilab-Pub-96/390-E

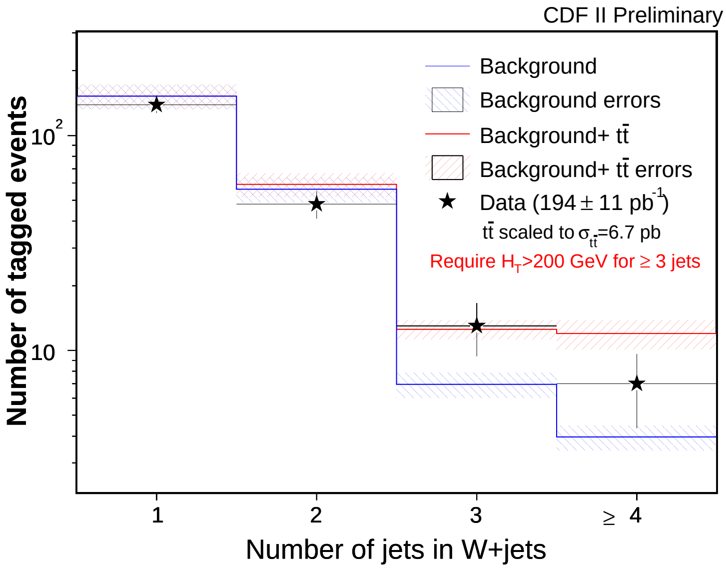

FIG. 6: The expected total background plus $t\bar{t}$ and observed tags in $W+$ 1, 2, 3 and 4 or more jet events. The background is corrected for the $t\bar{t}$ content of the pretagged sample. The $t\bar{t}$ expectation is normalized at 6.7pb.

[3] T. Sjostrand et al., High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. **135**, 238 (2001).
[4] D. Acosta, et al., Measurement of the $t\bar{t}$ Cross Section in $p\bar{p}$ Collisions at $\sqrt{s}$=1.96 TeV Using Lepton+Jets Events with Secondary Vertex b-tagging, To be submitted to Phys. Rev. D.
[5] G. Corcella et al., HERWIG 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (including supersymmetric processes), JHEP **01**, 10 (2001).